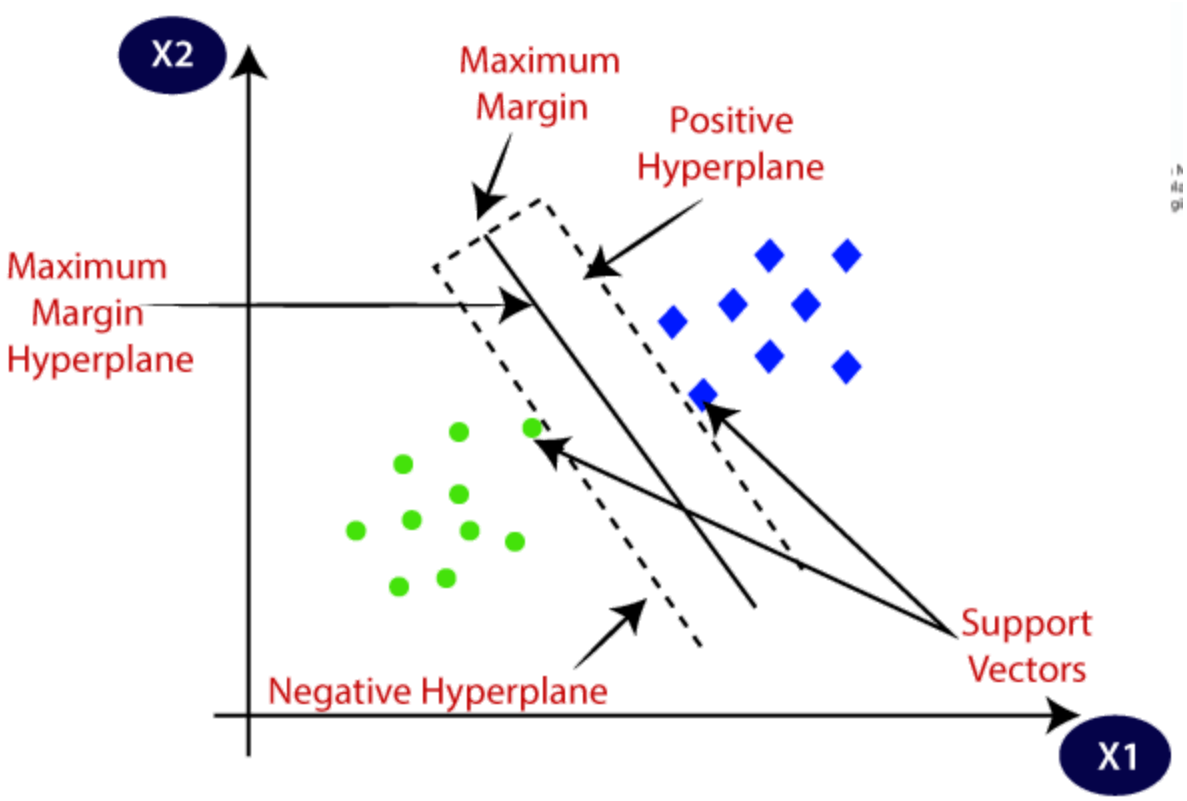
Brainstorming

1.Problem Statement:

Students are often worried about their chances of admission to University. The aim of this project is to help students in shortlisting universities with their profiles. The predicted output gives them a fair idea about their admission chances in a particular university. This analysis should also help students who are currently preparing or will be preparing to get a better idea.

2.Ideas:

2.1. Aarthi: We can predict the admission chances by using machine learning.We can use SVM algorithm to predict the admission condition.Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.



2.2. Ashlin: We can train our model and make it to be executed to all kinds of scenarios to achieve that we can go for Naive Bayes algorithm.Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems.It is mainly used in text classification that includes a high-dimensional training dataset. Naïve Bayes Classifier is one of the simple and most effective classification algorithms which helps in building the fast machine learning models that can make quick predictions.It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.Some popular examples of Naïve Bayes Algorithm are spam filtration, Sentimental analysis, and classifying articles.

Bayes' Theorem

- Bayes' theorem is also known as Bayes' Rule or Bayes' law, which is used to determine the probability of a hypothesis with prior knowledge. It depends on the conditional probability.
- The formula for Bayes' theorem is given as:

$$P(A|B)= \frac{P(B|A)P(A)}{P(B)}$$

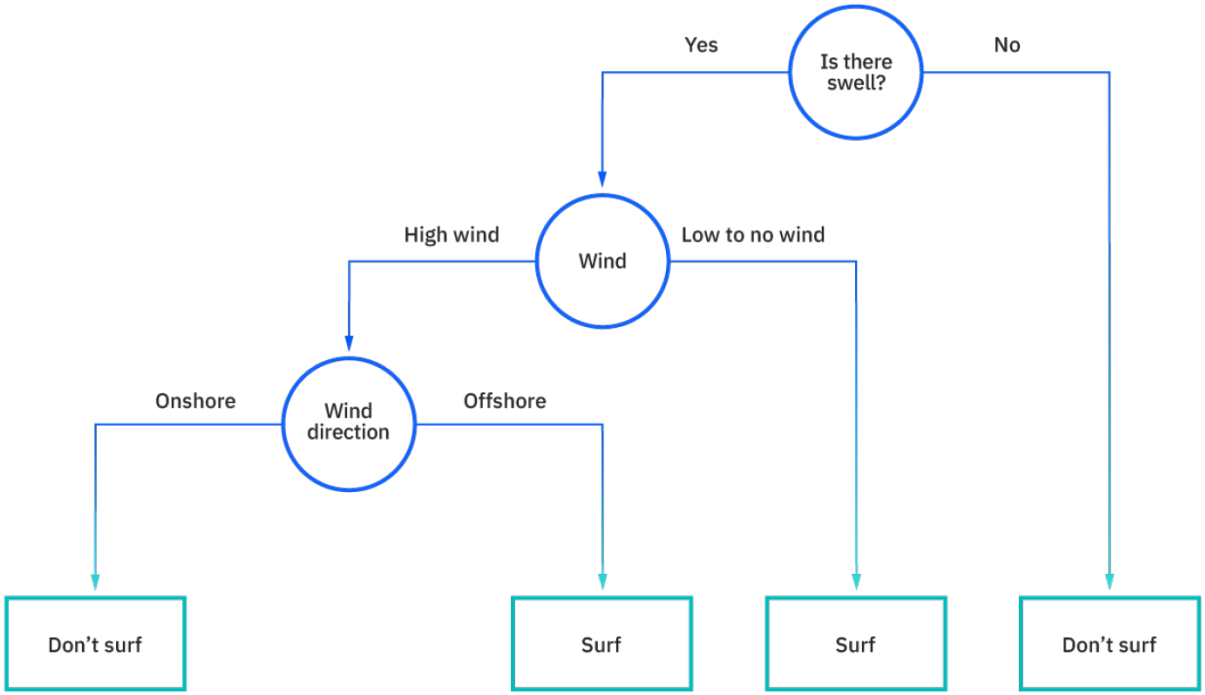
Where,

P(A|B) is Posterior probability: Probability of hypothesis A on the observed event B.

P(B|A) is Likelihood probability: Probability of the evidence given that the probability of a hypothesis is true.

2.3. Lekhashree: We can use the Decision tree algorithm to train our model.A decision tree is a non-parametric supervised learning algorithm, which is utilized for both classification and regression tasks. It has a hierarchical tree structure, which consists of a root node, branches, internal nodes and leaf nodes. A decision tree starts with a root node, which does not have any incoming branches. The outgoing branches from the root node then feed into the internal nodes, also known as decision nodes. Based on the available features, both node types conduct evaluations to form homogenous subsets, which are denoted by leaf nodes, or terminal nodes. The leaf nodes represent all

the possible outcomes within the dataset. As an example, let's imagine that you were trying to assess whether or not you should go surf, you may use the following decision rules to make a choice:



2.4. Harini: To obtain the predicted results we can use K-means algorithm.K-Means Clustering is an unsupervised learning algorithm that is used to solve the clustering problems in machine learning or data science.

Working of K-Means algorithm

Step-1: Select the number K to decide the number of clusters.

Step-2: Select random K points or centroids. (It can be different from the input dataset).

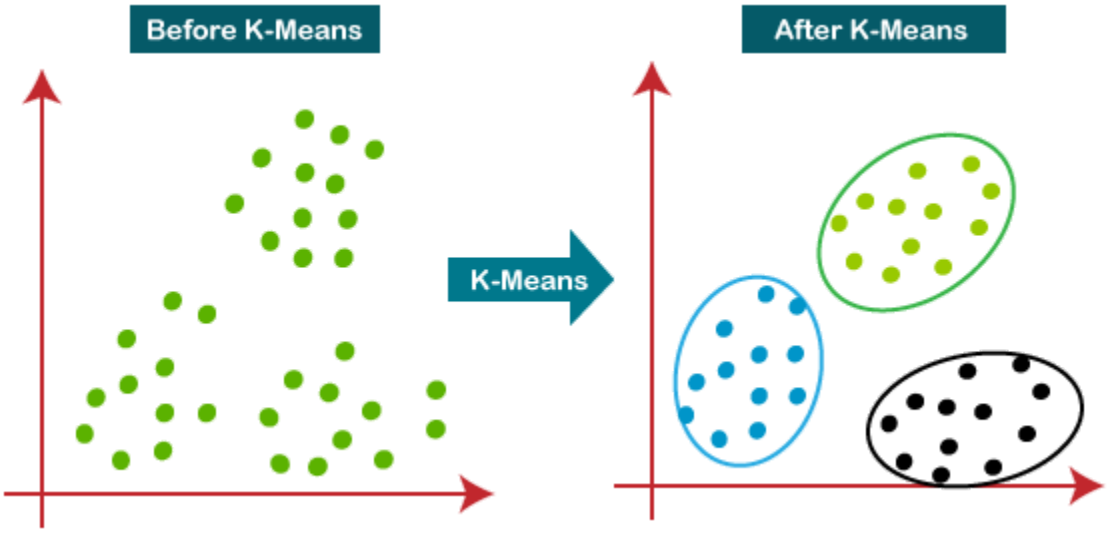
Step-3: Assign each data point to their closest centroid, which will form the predefined K clusters.

Step-4: Calculate the variance and place a new centroid of each cluster.

Step-5: Repeat the third steps, which means re-assign each datapoint to the new closest centroid of each cluster.

Step-6: If any reassignment occurs, then go to step-4 else go to FINISH.

Step-7: The model is ready.



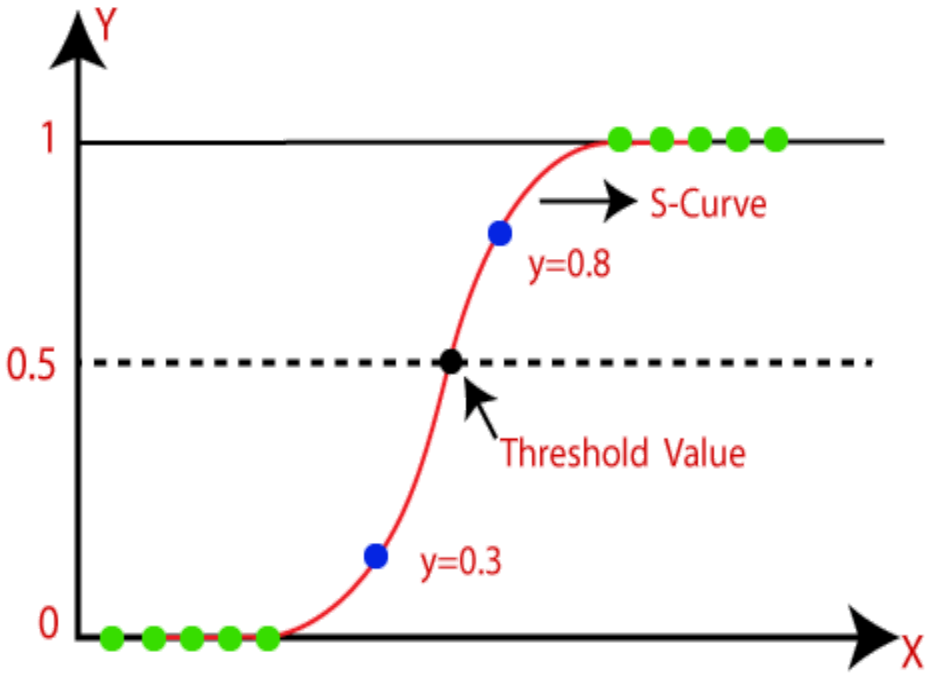
2.5. Monisha : To find the best and optimal solution to the selected problem statement that is the university admission predictor we can give the solution using machine learning. In machine learning there are lots of algorithms to predict the data and to give the output based on it's prediction.This prediction must be done in a correct manner and the

output must be generated as per the correct prediction.The result should have a high accuracy.To achieve that we can use logistic regression algorithm which have high accuracy almost equivalent to 85%.So from my side I have a confident on logistic regression algorithm.

Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, True or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.

Logistic Regression is much similar to Linear Regression except that how they are used. Linear Regression is used for solving Regression problems, whereas Logistic regression is used for solving the classification problems.Logistic Regression is a significant machine learning algorithm because it has the ability to provide probabilities and classify new data using continuous and discrete datasets.

Logistic Regression can be used to classify the observations using different types of data and can easily determine the most effective variables used for the classification. The below image is showing the logistic function:



Assumptions for Logistic Regression:

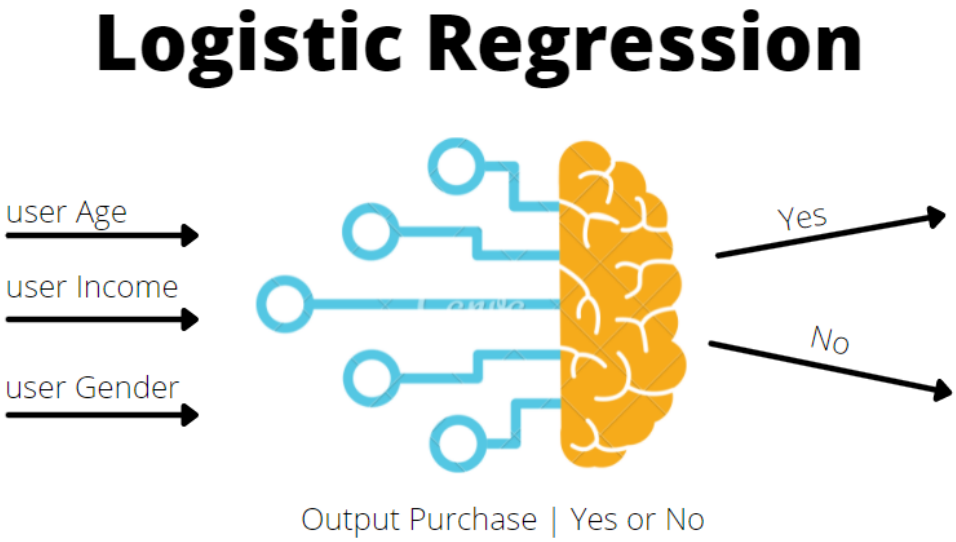
- The dependent variable must be categorical in nature.
- The independent variable should not have multicollinearity

Logistic Regression Equation:

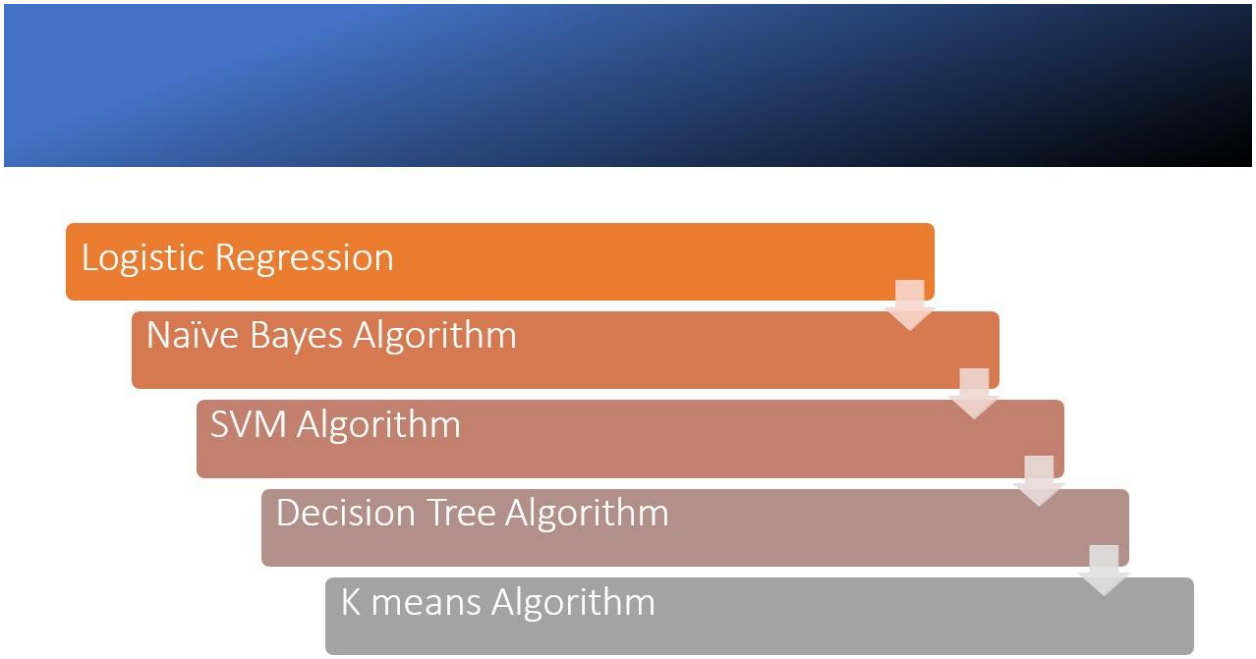
$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_nx_n$$

$\frac{y}{1-y}$; 0 for y= 0, and infinity for y=1

$$log\left[\frac{y}{1-y}\right] = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_nx_n$$



3.Idea Prioritization:



From this prioritization from top to bottom we decided to go with a Logistic regression algorithm to find the accurate output from our model.Because it has the highest accuracy compared with all other algorithms.It will give most accurate results among all other algorithms.

THANK YOU